# Hankel Low-Rank Approximation for Seismic Noise Attenuation

Chong Wang, Zhihui Zhu, *Student Member, IEEE*, Hanming Gu, Xinming Wu, and Shuaiqi Liu

***Abstract*— The low-rankness property of the Hankel matrix formulated from the clean seismic data corresponding to a few number of linear events has been successively leveraged in many low-rank (LR) approximation methods for seismic data denoising. The common scheme in these rank-reduction methods is to compute the best LR approximation of the formulated Hankel matrix and then obtain the denoised data from the LR matrix. However, without utilizing the Hankel structure when computing the LR approximation, if we rearrange the denoised data into a Hankel matrix, it is in general not exactly LR as expected. In this paper, we propose a Hankel LR (HLR) approximation method to simultaneously exploit both the Hankel structure and the LR property underlying the clean seismic data. The formulated HLR approximation problem is solved by an alternating-minimization-based algorithm. We provide rigorously convergence analysis of the proposed algorithm. The superior performance of the proposed HLR approximation method is demonstrated on both synthetic and field seismic data.**

***Index Terms*— Alternating minimization, Hankel matrix, low rank (LR), seismic noise attenuation.**

## I. Introduction

IT IS inevitable that random noise—which will influence the stability and precision of inversion-based migration and introduce artifacts in the migrated image—appears widely through the acquisition of seismic data [1]–[3]. Thus, one of the most important steps in seismic data processing is noise suppression [1], as removing the additive noise can facilitate the following steps, such as fault detection [4]–[6], seismic attributes extraction techniques [7], [8], and even for oil and gas detection [2], [3].

In the last two decades, a wide range of methods have been utilized to remove the additive noise and thus to improve the performance of the following seismic data processing tasks (such as poststack inversion). The fundamental idea underlying most of the denoising methods is to exploit the concise structure or property that is only within the true (also known as clean) signals, rather than the noise. In seeking to exploit this useful structure or model the true signals obey, one line of research attempts to represent the useful signals with some well-selected basis [9], [10], i.e., transform the data into different domains in which the true signals present very benign structure such as sparsity (i.e., only very few number of coefficients are large, which is a key ingredient in compressive sensing [11], [12]), which is not the case for the noise. Typical examples include Fourier transform [13], wavelet transform [14], curvelet transform [15], [16], shearlet transform [17], [18], seislet transform [19], [20], Radon transform [21], learned dictionary [22]–[24], and the deep learning technique [25]–[27]. The predictable property of the true signal has also been leveraged to enhance useful signals and suppress random noise in $t-x$ predictive filtering [28], $f-x$ deconvolution [29], the polynomial fitting-based approach [30], and nonstationary predictive filtering [31].

A second line of research attempts to directly extract the useful information. A natural way is to decompose the noisy seismic data into different components and then select the principal components as a surrogate of the true signal. Typical methods belonging to these decomposition-based denoising approaches include empirical-mode decomposition (EMD) and its improvements [32]–[35], singular value decomposition (SVD)-based approaches [36], [37], and regularized nonstationary decomposition [38], [39]. Another closely related method is the so-called low-rank (LR) approximation, such as the Cadzow filtering [40], singular spectrum analysis (SSA) [41], multichannel SSA [42], and damped SSA [2], [3], [43].

Among these methods, the LR approximation approaches have been widely adopted in the field of random noise suppression due to its outstanding performance, especially in land seismic surveys where most subsurface reflectors are relatively flat [3], [42], [44]. The LR model has also been successfully exploited for reverse time migration [45]. The theoretical underpinning of the LR-based methods [41], [42], [44] is that the Hankel matrix formulated from the useful seismic signal is LR and the *rank* is equal to the number of linear events (or dip components) in the seismic profile (see Section II for the details), while the Hankel matrix corresponding to the noise component is in general high rank. To exploit such an LR property, the general step in these LR approximation

Manuscript received April 3, 2018; revised June 5, 2018; accepted July 16, 2018. This work was supported in part by the Important Science and Technology Specific Projects under Grant P17057-9 and in part by Natural Science Foundation of China under Grant 61401308 and Grant 61572063. *(Corresponding author: Zhihui Zhu.)*

C. Wang is with the Institute of Geophysics and Geomatics, China University of Geosciences, Wuhan 430074, China, and also with the Bureau of Economic Geology, The University of Texas at Austin, Austin, TX 78713 USA (e-mail: chongwang@cug.edu.cn).

Z. Zhu is with the Center of Imaging Science, Johns Hopkins University, Baltimore, MD 21218 USA (e-mail: zzhu29@jhu.edu).

H. Gu is with the Institute of Geophysics and Geomatics, China University of Geosciences, Wuhan 430074, China (e-mail: hmgu@cug.edu.cn).

X. Wu is with the Bureau of Economic Geology, The University of Texas at Austin, Austin, TX 78713 USA (e-mail: xinming.wu@beg.utexas.edu).

S. Liu is with the College of Electronic and Information Engineering, Hebei University, Baoding 071002, China (e-mail: shdkj-1918@163.com).

Color versions of one or more of the figures in this paper are available online at http://ieeexplore.ieee.org.

Digital Object Identifier 10.1109/TGRS.2018.2858545

methods is to rearrange the noisy data $\boldsymbol{d} = \boldsymbol{s} + \boldsymbol{n}$ (where $\boldsymbol{s}$ is the clean signal and $\boldsymbol{n}$ is the additive noise) into a Hankel matrix $\boldsymbol{D}$, compute the LR approximation $\boldsymbol{L}$ of this Hankel matrix $\boldsymbol{D}$, and finally obtain the estimation $\widehat{\boldsymbol{s}}$ by transferring $\boldsymbol{L}$ back to the signal (i.e., taking the average of the skew diagonals of $\boldsymbol{L}$). The last step is equivalent to first compute a Hankel matrix $\boldsymbol{H}$ that is closest to this LR approximation $\boldsymbol{L}$ and then map this Hankel matrix $\boldsymbol{H}$ back to the signals' domain (the inverse of mapping a signal to a Hankel matrix). However, we note that the LR approximation $\boldsymbol{L}$ is, in general, no longer to be a Hankel matrix. On the other hand, when we project this LR approximation $\boldsymbol{L}$ onto the Hankel space, the projection $\boldsymbol{H}$ actually has different (higher) ranks than $\boldsymbol{L}$. In other words, if we formulate a Hankel matrix from the estimated signal $\widehat{\boldsymbol{s}}$, this Hankel matrix is, in general, not exactly LR as expected (see Fig. 1 and the detailed discussion in Section III-A). This suggests an interesting fact that by exploiting the additional Hankel structure when computing the LR approximation, the performance of LR approximation methods can be further improved.

Based on the above-mentioned observation, in this paper, we propose a Hankel LR (HLR) approximation method to exploit the additional Hankel structure in computing the LR approximation. The contribution of this paper is summarized as follows.

The first contribution of this paper is to simultaneously exploit both the Hankel structure and the LR property underlying the useful seismic data. In particular, we formulate the denoising problem as an HLR approximation rather than the simple LR approximation adopted in the conventional methods [3], [42], [44]. By utilizing the additional Hankel structure, which actually has a very few number of degrees of freedom (DOFs), we expect to improve the denoising performance. We also propose an alternating-minimization-based algorithm solving the formulated HLR approximation problem.

The second contribution is to provide rigorously convergence analysis of the proposed algorithm. In particular, we show that the algorithm has subsequence convergence and any of its limit point is a coordinatewise minimum point and a stationary point.

The outline of this paper is as follows. Section II contains the fundamental material explaining the low-rankness property of the Hankel matrix formulated from the useful seismic data. We propose the HLR approximation method for seismic denoising in Section III. In Section IV, we demonstrate the performance of the proposed method on both synthetic seismic data and field seismic data. We conclude this paper in Section V.

## II. Preliminaries

Consider 2-D seismic data acquired on a regular grid $D(\ell, w)$, where $\ell = 1, \ldots, L$ denotes trace indices and $w$ denotes the temporal frequency. According to [42], a linear event has a space–frequency representation given by $\sigma(w)e^{-jw\ell p}$, where $p$ represents the slowness and $\sigma(w)$ denotes the amplitude. Considering a small window of analysis that consists of a superposition of $R$ (which is much smaller than $L$) linear events, we have

$$D(\ell, w) = S(\ell, w) + N(\ell, w) \tag{1}$$

where $N(\ell, w)$ is the noise component and

$$S(\ell, w) = \sum_{r=1}^{R} \sigma_r(w) e^{-jw\ell p_r} \tag{2}$$

corresponds to the signal component. For convenience, for each frequency $w$, we stack $D(1, w), D(2, w), \ldots, D(L, w)$ into a length-$L$ vector

$$\boldsymbol{d}_w = [D(1, w) \quad D(2, w) \quad \cdots \quad D(L, w)]^{\mathrm{T}}$$

where T represents the transpose operator. Similarly

$$\begin{aligned} \boldsymbol{s}_w &= [S(1, w) \quad S(2, w) \quad \cdots \quad S(L, w)]^{\mathrm{T}}, \\ \boldsymbol{n}_w &= [N(1, w) \quad N(2, w) \quad \cdots \quad N(L, w)]^{\mathrm{T}}. \end{aligned}$$

From (1), we have

$$\boldsymbol{d}_w = \boldsymbol{s}_w + \boldsymbol{n}_w.$$

To utilize the underlying structure within the seismic data $S(\ell, w)$ in (2), a popular approach is to map the vector $\boldsymbol{d}_w$ into a Hankel matrix. Toward that end, we first define an operator $\mathcal{H} : \mathbb{R}^L \to \mathbb{R}^{M\times N}$ that maps a vector into a Hankel matrix

$$\mathcal{H}(\boldsymbol{x}) := \begin{bmatrix} x_1 & x_2 & x_3 & \cdots & x_N \\ x_2 & x_3 & \cdots & \cdots & x_{N+1} \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ x_M & x_{M+1} & \cdots & \cdots & x_L \end{bmatrix} \tag{3}$$

where $x_\ell$ is the $\ell$th element in $\boldsymbol{x}$, and $M$ and $N$ are usually chosen as $M = \lfloor L/2 \rfloor + 1$ and $N = L - M + 1$ so that $\mathcal{H}(\boldsymbol{x})$ is close to a square matrix as possible [44]. Here, $\lfloor a \rfloor$ denotes the largest integer that is smaller than or equal to $a$. Throughout, the inverse Hankel operator is denoted by $\mathcal{H}^\dagger : \mathbb{R}^{M\times N} \to \mathbb{R}^L$, i.e., $\boldsymbol{x} = \mathcal{H}^\dagger(\boldsymbol{X})$ is obtained by taking the average of the skew diagonals of $\boldsymbol{X}$. Note that $\mathcal{H}^\dagger\mathcal{H} = \mathbf{I}$, but $\mathcal{H}\mathcal{H}^\dagger$ is not an identity operator.

Now for each frequency $w$, we map the vector $\boldsymbol{d}_w$ into a Hanekl matrix as

$$\boldsymbol{D}_w =: \mathcal{H}(\boldsymbol{d}_w) = \mathcal{H}(\boldsymbol{s}_w + \boldsymbol{n}_w) = \boldsymbol{S}_w + \boldsymbol{N}_w \tag{4}$$

where $\boldsymbol{S}_w = \mathcal{H}(\boldsymbol{s}_w)$ and $\boldsymbol{N}_w = \mathcal{H}(\boldsymbol{n}_w)$. To see the LR structure in $\boldsymbol{S}_w$, we rewrite

$$\begin{aligned} \boldsymbol{S}_w(m, n) &= \sum_{r=1}^{R} \sigma_r(w) e^{-jw\ (m+n-1)p_r} \\ &= \sum_{r=1}^{R} \sigma_k(w) e^{jwp_r} e^{-jwp_r m} e^{-jwp_r n} \end{aligned}$$

which implies

$$\boldsymbol{S}_w = \sum_{r=1}^{R} \sigma_r(w) e^{jwp_r} \boldsymbol{e}_{f_r,M} \boldsymbol{e}_{f_r,N}^{\mathrm{T}} \tag{5}$$

with

$$\boldsymbol{e}_{w,p_r,M} = [e^{-jwp_r} \quad e^{-jwp_r 2} \quad \cdots \quad e^{-jwp_r M}]^{\mathrm{T}}.$$

Thus for each frequency $w$, $\boldsymbol{S}_w$ has rank at most $R$, as it can be expressed in (5) as the sum of $R$ rank-1 matrices, no matter what the values of $p_r$ are for all $r = 1, \ldots, R$. We stress that this lies out the main difference between the LR approach and the Fourier transform-based denoising; when the frequencies $p_r$ are on the discrete Fourier transform (DFT) grid, the DFT coefficients of $\boldsymbol{s}_w$ are sparse and one can utilize this sparsness to remove the noise from $\boldsymbol{d}_w$. However, when the frequencies $p_r$ are not on the discrete DFT grid, the DFT coefficients of $\boldsymbol{s}_w$ are not sparse, which are known as the "leakage" issue [10]. On the other hand, the LR approach avoids this "leakage" issue, since the Hankel matrix $\boldsymbol{S}_w$ is LR for any values of $p_r$. This explains the potential advantages of rank-reduction methods compared with the Fourier transform-based approaches for seismic noise denoising.

### A. Seismic Low-Rank Approximation

The LR structure in $\boldsymbol{S}_w$ has been popularly utilized for attenuating random noise in seismic data [42], [44], where for each $w$, the LR Hankel matrix $\boldsymbol{S}_w$ is estimated by

$$\overline{\boldsymbol{S}}_w = \underset{\text{rank}(\boldsymbol{S})\leq R}{\arg\min} \|\boldsymbol{D}_w - \boldsymbol{S}\|_F^2 \tag{6}$$

which is a rank-$R$ approximation to the Hankel matrix $\boldsymbol{D}_w$ formulated from the data $\boldsymbol{d}_w$. It follows from the Eckart–Yound theorem that the optimal solution to (6) is given by:

$$\overline{\boldsymbol{S}}_w = [\boldsymbol{S}_w]_R := \boldsymbol{U}_R \boldsymbol{\Sigma}_R \boldsymbol{V}_R^{\mathrm{H}} \tag{7}$$

where $\boldsymbol{\Sigma}_R$ is an $R \times R$ diagonal matrix consisting the first $R$ largest singular values and $\boldsymbol{U}_R$ and $\boldsymbol{V}_R$ containing the corresponding $R$ right singular vectors and $R$ left singular vectors, respectively. The best rank-$R$ approximation is the key ingredient for filtering or denoising [42], [44].

## III. Seismic Hankel Low-Rank Approximation

In this section, we provide an HLR approximation for attenuating random noise in seismic data. Unlike the LR approximation method (in Section II-A) which only exploits the LR structure in $\boldsymbol{S}_w$, the proposed HLR approximation method simultaneously utilizes the LR structure and the Hankel structure in $\boldsymbol{S}_w$. By exploiting the additional Hanekl structure in $\boldsymbol{S}_w$, the proposed method is expected to have a better performance in removing noise. To simplify the notations and better present the main idea, in this section, we drop the subscript $w$ in $\boldsymbol{D}_w$, $\boldsymbol{S}_w$, and $\boldsymbol{N}_w$ and write them as $\boldsymbol{D}$, $\boldsymbol{S}$, and $\boldsymbol{N}$, respectively. That is

$$\boldsymbol{D} =: \mathcal{H}(\boldsymbol{d}) = \mathcal{H}(\boldsymbol{s} + \boldsymbol{n}) = \boldsymbol{S} + \boldsymbol{N} \tag{8}$$

where $\boldsymbol{S} = \mathcal{H}(\boldsymbol{s})$, $\boldsymbol{N} = \mathcal{H}(\boldsymbol{n})$, $\boldsymbol{d}$ is the measurement vector, $\boldsymbol{n}$ is the noise vector, and $\boldsymbol{s}$ is the signal vector that needs to be estimated.

### A. Importance of Hankel Structure

To intuitively explain the importance of exploiting the Hankel structure within $\boldsymbol{S}$ when we remove the noise component $\boldsymbol{N}$ from $\boldsymbol{D}$, we consider the number of degrees of freedom (DOF) within the Hankel structure and the LR structure. For a general Hankel matrix $\boldsymbol{S} \in \mathbb{R}^{M\times N}$, its number of DOF is $M + N$. On the other hand, for a general rank-$R$ $\boldsymbol{S} \in \mathbb{R}^{M\times N}$, the number of DOF is $R(M + N - R)$, which in general is much higher than $(M + N)$, especially when $R \gg 1$. As we explained before, the fundamental idea underlying most of denoising methods is to exploit the concise structure that the signals obey. To utilize the concise structure within a signal is equivalent to come up with a method to leverage its small number of DOF within the signal. In this sense, the Hankel structure within $\boldsymbol{S}$ is at least as useful as the low-rankness property of $\boldsymbol{S}$.

We now utilize a simple example shown in Fig. 1 to further explain the potential benefit by exploiting both the Hankel structure and the low-rankness within $\boldsymbol{S}$. We generate a rank-2 Hankel matrix [see Fig. 1(a1)], whose singular values are shown in Fig. 1(a2). According to (8), we then generate a random Gaussian noise $\boldsymbol{n}$ and add the corresponding Hankel noise $\mathcal{H}(\boldsymbol{n})$ with $\boldsymbol{S}$ to get the noisy Hankel matrix $\boldsymbol{D}$, which is shown in Fig. 1(b1). As observed from Fig. 1(b1), most of the singular values of $\boldsymbol{D}$ are large, indicating that $\boldsymbol{D}$ is not LR. Based on (6), we compute the rank-2 approximation $\overline{\boldsymbol{S}}$ of $\boldsymbol{D}$ and show it in Fig. 1(c1). Even though $\overline{\boldsymbol{S}}$ has the same rank as $\boldsymbol{S}$, the former is not a Hankel matrix and is not close to the later. Fig. 1(d1) shows $\mathcal{H}(\mathcal{H}^{\dagger}(\overline{\boldsymbol{S}}))$, the projection of $\overline{\boldsymbol{S}}$ into the Hankel space. It is clear from Fig. 1(d2) that this Hankel matrix is not LR. We note that this classical procedure [i.e., obtain the LR approximation $\overline{\boldsymbol{S}}$ and $\mathcal{H}^{\dagger}(\overline{\boldsymbol{S}})$] is utilized in [2], [3], [41], and [42].

To overcome the above issue, we exploit the additional Hankel structure when we compute an LR approximation to $\boldsymbol{D}$. To that end, we apply the proposed HLR approximation (see Algorithm 1) to obtain an LR approximation $\boldsymbol{L}$, which is also close to a Hankel matrix. Fig. 1(e1) shows the obtained $\boldsymbol{L}$. Compared to $\overline{\boldsymbol{S}}$ in Fig. 1(c2), $\boldsymbol{L}$ is a better estimation for the true Hanekl matrix $\boldsymbol{S}$.

### B. Hankel Low-Rank Approximation for Seismic Donoising

Inspired by the above discussion, we now formulate the HLR seismic denoising as

$$\begin{aligned} &\min_{\boldsymbol{H}\in\mathbb{R}^{M\times N}} \|\boldsymbol{D} - \boldsymbol{H}\|_F^2 \\ &\text{s.t. } \text{rank}(\boldsymbol{H}) \leq R, \ \ \boldsymbol{H} \in \mathbb{H} \end{aligned} \tag{9}$$

where

$$\mathbb{H} = \{\boldsymbol{H} \in \mathbb{R}^{M\times N} : \boldsymbol{H} = \mathcal{H}\mathcal{H}^{\dagger}(\boldsymbol{H})\} \tag{10}$$

is the set of $M \times N$ Hankel matrices. Here, $\mathcal{H}$ is the operator defined in (3). Without the Hankel constraint, (9) reduces to the conventional LR approximation for seismic denoising. By exploiting this underlying Hankel structure within $\boldsymbol{S}$, we expect to obtain a better estimation of $\boldsymbol{S}$. However, unlike (6) whose closed-form solution can be obtained by SVD, in general there is no such closed-form solution for (9) because of the additional Hankel structure.

To develop an algorithm for solving (9), it is useful to introduce an auxiliary variable $\boldsymbol{L} \approx \boldsymbol{H}$ and consider a slight

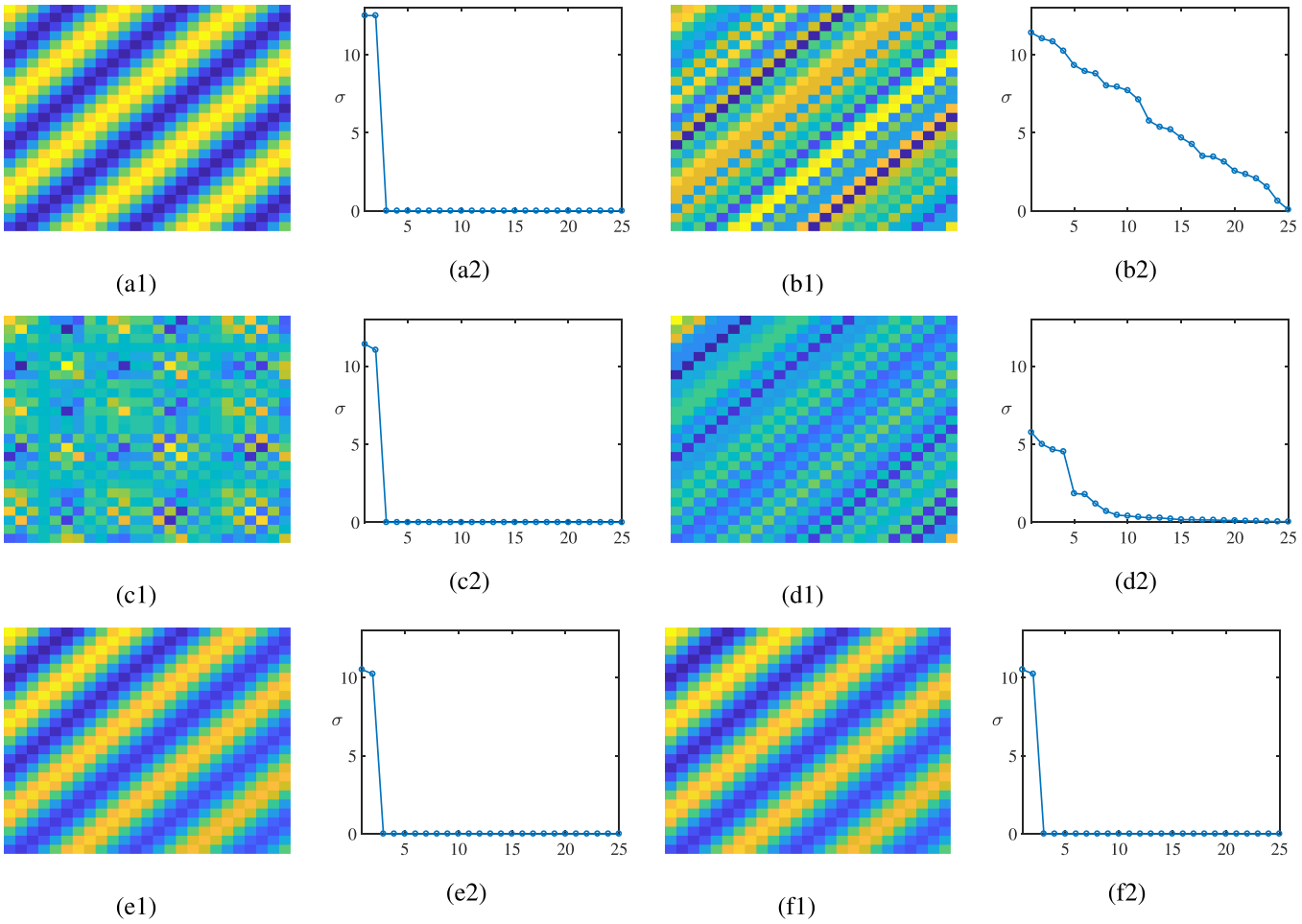


Fig. 1. (a1) Rank-2 Hankel matrix $\boldsymbol{S}$. (a2) Singular values of $\boldsymbol{S}$. (b1) Noised Hankel matrix $\boldsymbol{D} = \boldsymbol{S} + \mathcal{H}(\boldsymbol{n})$, where $\boldsymbol{n}$ is the random noise. (b2) Singular values of $\boldsymbol{D}$. (c1) $\overline{\boldsymbol{S}}$ that is a rank-2 approximation of $\boldsymbol{D}$. (c2) Singular values of $\overline{\boldsymbol{S}}$. (d1) Hankel matrix $\mathcal{H}(\mathcal{H}^{\dagger}(\overline{\boldsymbol{S}}))$ corresponding to $\overline{\boldsymbol{S}}$. (d2) Singular values of $\mathcal{H}(\mathcal{H}^{\dagger}(\overline{\boldsymbol{S}}))$. (e1) LR matrix $\boldsymbol{L}$ obtained by Algorithm 1. (e2) Singular values of $\boldsymbol{L}$. (f1) Hankel matrix $\boldsymbol{H}$ obtained by Algorithm 1. (f2) Singular values of $\boldsymbol{H}$.

relaxation of (9)

$$\begin{aligned} \min_{\boldsymbol{H},\boldsymbol{L}\in\mathbb{R}^{M\times N}} & f(\boldsymbol{H},\boldsymbol{L}) := \|\boldsymbol{D}-\boldsymbol{H}\|_F^2 + \lambda\|\boldsymbol{H}-\boldsymbol{L}\|_F^2 \\ \text{s.t. } & \operatorname{rank}(\boldsymbol{L}) \le R,\ \ \boldsymbol{H}\in\mathbb{H} \end{aligned} \tag{11}$$

where $\lambda$ is a penalty parameter. We now utilize the alternating minimization to solve (11). An alternating minimization method has been widely utilized for solving practical problem [46], [47]. This method starts with an initial $\boldsymbol{L}_0$ and alternatively optimizes with respect to $\boldsymbol{H}$ (when $\boldsymbol{L}$ is fixed) and with respect to $\boldsymbol{L}$ (when $\boldsymbol{H}$ is fixed)

$$\boldsymbol{H}_{k+1} = \arg\min_{\boldsymbol{H}\in\mathbb{H}} f(\boldsymbol{H},\boldsymbol{L}_k) \tag{12}$$

$$\boldsymbol{L}_{k+1} = \arg\min_{\operatorname{rank}(\boldsymbol{L})\le R} f(\boldsymbol{H}_{k+1},\boldsymbol{L}) \tag{13}$$

where $\boldsymbol{H}_k$ and $\boldsymbol{L}_k$ denote the values of $\boldsymbol{H}$ and $\boldsymbol{L}$ in the $k$th iteration, respectively. To solve (12), we first rewrite $f(\boldsymbol{H},\boldsymbol{L}_k)$ by

$$\begin{aligned} f(\boldsymbol{H},\boldsymbol{L}_k) &= \|\boldsymbol{D}-\boldsymbol{H}\|_F^2 + \lambda\|\boldsymbol{H}-\boldsymbol{L}_k\|_F^2 \\ &= (1+\lambda)\left\|\boldsymbol{H} - \frac{1}{1+\lambda}(\boldsymbol{D}+\lambda\boldsymbol{L}_k)\right\|_F^2 + c \end{aligned}$$

where $c$ is independent of $\boldsymbol{H}$. Thus, (12) has a simple closed-form solution

$$\boldsymbol{H}_{k+1} = \mathcal{P}_{\mathbb{H}}\left(\frac{1}{1+\lambda}(\boldsymbol{D}+\lambda\boldsymbol{L}_k)\right) = \frac{1}{1+\lambda}\mathcal{P}_{\mathbb{H}}(\boldsymbol{D}+\lambda\boldsymbol{L}_k)$$

where $\mathcal{P}_{\mathbb{H}} : \mathbb{R}^{M\times\mathbb{N}} \to \mathbb{H}$ is the orthogonal projector onto the set of Hankel matrices $\mathbb{H}$. In particular, we have $\mathcal{P}_{\mathbb{H}}(\cdot) = \mathcal{H}^{\dagger}(\mathcal{H}(\cdot))$. Similarly, there also exists a simple closed-form solution for (13)

$$\boldsymbol{L}_{k+1} = \mathcal{P}_{\mathbb{L}}(\boldsymbol{H}_{k+1})$$

where $\mathcal{P}_{\mathbb{L}}$ is the projection onto the set of $M \times N$ LR matrices

$$\mathbb{L} := \{\boldsymbol{L}\in\mathbb{R}^{M\times N} : \operatorname{rank}(\boldsymbol{L}) \le R\}. \tag{14}$$

In other words, $\mathcal{P}_{\mathbb{L}}(\boldsymbol{A})$ computes the best rank-$R$ approximation (see (7)) of $\boldsymbol{A}$. The full procedure is depicted in Algorithm 1. The following result establishes the convergence guarantee for Algorithm 1.

*Theorem 1 (Convergence of Algorithm 1):* Let $(\boldsymbol{H}_k,\boldsymbol{L}_k)$ be the sequence generated by Algorithm 1. Then, the following holds.

1) The sequence $\{\boldsymbol{H}_k\}$ is regular
$$\lim_{k\to\infty} \|\boldsymbol{H}_k - \boldsymbol{H}_{k+1}\|_F = 0. \tag{15}$$
2) The sequence $(\boldsymbol{H}_k, \boldsymbol{L}_k)$ is a bounded sequence and thus it has at least one convergent subsequence. For any convergent subsequence with limit point $(\boldsymbol{H}^\star, \boldsymbol{L}^\star)$, we have $\boldsymbol{H}^\star \in \mathbb{H}$, $\mathbb{L}^\star \in \mathbb{L}$ and
$$f(\boldsymbol{H}^\star, \boldsymbol{L}^\star) = \inf_k f(\boldsymbol{H}_k, \boldsymbol{L}_k).$$
3) Furthermore, $(\boldsymbol{H}^\star, \boldsymbol{L}^\star)$ is not only a coordinatewise minimum point but also a stationary point of (11).

The proof of Theorem 1 is in Appendix A. In other words, Theorem 1 assures that Algorithm 1 finds a stationary point of (11) and the iterates of $\{\boldsymbol{H}_k\}$ are regular and hence we can obtain the denoise signal by applying $\mathcal{H}^\dagger$ to the last iterate of $\boldsymbol{H}$. By utilizing the penalty $\|\boldsymbol{H} - \boldsymbol{L}\|_F^2$ in (11), $\boldsymbol{H}^\star$ and $\boldsymbol{L}^\star$ are supposed to be close to each other, and thus the estimation obtained from $\boldsymbol{H}^\star$ is expected to be a better estimator of the clean signal compared to the one obtained by the classical LR approximation method.

With respect to the computational complexity of Algorithm 1, in each iteration, it has two steps. They are projection onto the Hankel space (i.e., updating $\boldsymbol{H}_k$), which is very cheap and requires $O(MN)$ operations, and LR approximation (i.e., updating $\boldsymbol{L}_k$), which requires $O(M^2N)$ operations. Thus, the total cost for Theorem 1 is $O(M^2NN_{ite})$. As a contrast, the damped LR (DLR) approximation method [3] and the classical LR approximation method [42] require performing SVD, and the cost is $O(M^2N)$. Fortunately, as demonstrated in Section IV with experiments, Algorithm 1 converges very fast and thus we can choose a very small $N_{ite}$ (say $N_{ite} = 4$).

We end this section by discussing the selection of the rank $R$. As we explained before in Section II, the Hankel matrix $\boldsymbol{S} = \mathcal{H}(\boldsymbol{s})$ is LR only when the signal $\boldsymbol{s}$ corresponds to $R$ linear events. However, the real seismic data are in general very complicated and are generated from nonlinear events. Thus, a common strategy is to work in local windows, since, within a small window, the vent can be approximately viewed as a superposition of a few number of linear events [3], [42], [44]. However, this poses the difficulty of choosing an appropriate rank for each local processing window. In other words, to optimize the performance of the HLR approximation for local seismic denoising, we need to choose the optimal rank for each window. To overcome this issue, Chen *et al.* [44] utilized the EMD by decomposing the signal into a number of components corresponding to linear events. In this paper, we utilize a more straightforward way to adaptively select the rank for each Hankel matrix. In particular, suppose $\boldsymbol{Y} = \boldsymbol{X} + \boldsymbol{N}$, where $\boldsymbol{N} \in \mathbb{R}^{M\times N}$ (with $M \le N$) is a Gaussian random matrix and $\boldsymbol{X}$ is an LR matrix but its true rank $R^\star$ is unknown. Let $\sigma_1 \ge \cdots \ge \sigma_M$ be the singular values of $\boldsymbol{Y}$. Without knowing the statistical information (like the variance) of the noise $\boldsymbol{N}$, Gavish and Donoho [48] provided a simple rule for estimating the rank of $\boldsymbol{X}$
$$R = \#\{i : \sigma_i \ge c\sigma_{\text{med}}\} \tag{16}$$
where # denotes the number of elements in a set, $c = 0.56(M/N)^3 - 0.95(M/N)^2 + 1.82(M/N) + 1.43$, and $\sigma_{\text{med}}$ is the median singular value of the matrix $\boldsymbol{Y}$. In other words, $R$ in (16) is equal to the number of singular values that are greater than $c\sigma_{\text{med}}$. The computational cost for (16) is negligible compared to the cost for SVD. We incorporate (16) into Algorithm 1 to adaptively estimate the rank $R$. However, we note that it is possible to obtain a better estimation by exploiting the Hankel structure in the noise matrix $\boldsymbol{N}$ in (8). We defer it to the future work.

**Algorithm 1** HLR Approximation for Seismic Donoising

**Input:** noisy data $\boldsymbol{d}$, rank $R$, number of iterations $N_{ite}$
**Formulating a Hankel matrix:** $\boldsymbol{D} = \mathcal{H}(\boldsymbol{d})$
**Initialization:** $\boldsymbol{L}_0 = \mathcal{P}_{\mathbb{L}}(\boldsymbol{D})$
**for** $k = 0, 1, \ldots, N_{ite} - 1$ **do**
$$\boldsymbol{H}_{k+1} = \frac{1}{1+\lambda}\mathcal{P}_{\mathbb{H}}(\boldsymbol{D} + \lambda\boldsymbol{L}_k)$$
$$\boldsymbol{L}_{k+1} = \mathcal{P}_{\mathbb{L}}(\boldsymbol{H}_{k+1})$$
**end for**
**Output:** the estimation of $\boldsymbol{s}$: $\mathcal{H}^\dagger(\boldsymbol{H}_{N_{ite}})$

## IV. Experiments

### A. Evaluation of Denoising Performance

To begin, we first give out some quantities to evaluate the denoising performance. For the synthetic examples, we use the following signal-to-noise ratio (SNR) [44]:
$$\text{SNR} = 10\log_{10}\frac{\|\boldsymbol{C}\|_F^2}{\|\boldsymbol{C} - \boldsymbol{A}\|_F^2} \tag{17}$$
where $\boldsymbol{C}$ is the clean seismic data and $\boldsymbol{A}$ is the estimated or denoised one.

For the field data example, since the true signal is unknown as a prior, we can compute the local similarity [44] or the local correlation coefficient to measure the difference between denoised data and the removed noise. The following are equations to compute the local correlation coefficient:
$$\rho_c = \frac{\langle\widehat{\boldsymbol{A}}, \widehat{\boldsymbol{B}}\rangle}{\|\widehat{\boldsymbol{A}}\|_F\|\widehat{\boldsymbol{B}}\|_F} \tag{18}$$
where $\widehat{\boldsymbol{A}}$ and $\widehat{\boldsymbol{B}}$ have elements
$$\hat{\boldsymbol{a}}_{ij} = \boldsymbol{a}_{ij} - \frac{1}{N_1N_2}\sum_{i=1}^{N_1}\sum_{j=1}^{N_2}\boldsymbol{a}_{ij} \tag{19}$$
$$\hat{\boldsymbol{b}}_{ij} = \boldsymbol{b}_{ij} - \frac{1}{N_1N_2}\sum_{i=1}^{N_1}\sum_{j=1}^{N_2}\boldsymbol{b}_{ij} \tag{20}$$
where $N_1$ and $N_2$ are the number of time-samples and traces in the local window, and $\boldsymbol{a}_{ij}$ and $\boldsymbol{b}_{ij}$ are the elements in the denoised data and the removed noise data, respectively. In other words, a high local correlation coefficient between denoised signal and noise indicates that useful signals exist

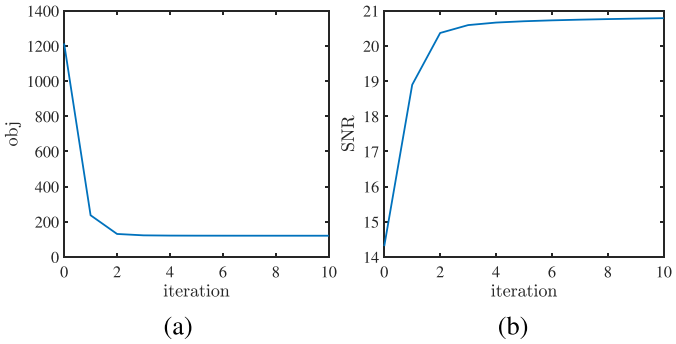


Fig. 2. Illustration of the convertence of Algorithm 1 in terms of (a) objective function $\|\boldsymbol{D}-\boldsymbol{H}_k\|_F^2+\lambda\|\boldsymbol{H}_k-\boldsymbol{L}_k\|_F^2$ as in (11) and (b) SNR: $10\log_{10}(\|\boldsymbol{S}\|_F^2)/(\|\boldsymbol{L}_k-\boldsymbol{S}\|_F^2)$.

in the removed noise, while a low local correlation coefficient (such as zero) between denoised signal and noise indicates superior denoising performance.

### *B. Convergence of Algorithm 1 for Hankel Low-Rank Approximation*

To demonstrate the converence of Algorithm 1, we generate a length-$L$ ($L=49$) and signal $\boldsymbol{s}$ consisting of $R=5$ linear events as $s_\ell=\sum_{r=1}^{R}\sigma_r e^{-j2\pi\ell p_r}$, where the frequencies $\{p_r\}$ are randomly chonsen from $[0,1]$ and the coefficients $\{\sigma_r\}$ are generated from a normal distribution. The measurement $\boldsymbol{d}\in\mathbb{R}^L$ is then generated by adding a Gaussian random noise $\boldsymbol{n}\in\mathbb{R}^L$, i.e., $\boldsymbol{d}=\boldsymbol{s}+\boldsymbol{n}$. We then formulate an $M\times N$ (with $M=N=25$) Hanekl matrix $\boldsymbol{D}=\mathcal{H}(\boldsymbol{d})=\mathcal{H}(\boldsymbol{s}+\boldsymbol{n})=\boldsymbol{S}+\boldsymbol{N}$, where $\boldsymbol{S}$ is a rank-$R$ Hankel matrix corresponding to the clean signal $\boldsymbol{s}$. We set $\lambda=10$ and perform Algorithm 1 to estimate $\boldsymbol{S}$ from $\boldsymbol{D}$. In Fig. 2, we display the objective function $\|\boldsymbol{D}-\boldsymbol{H}_k\|_F^2+\lambda\|\boldsymbol{H}_k-\boldsymbol{L}_k\|_F^2$ [see (11)] as well as the SNR of $\boldsymbol{L}_k$ [i.e., $10\log_{10}(\|\boldsymbol{S}\|_F^2)/(\|\boldsymbol{L}_k-\boldsymbol{S}\|_F^2)$] against the iteration, where $k$ denotes the $k$th iteration. As seen from Fig. 2, Algorithm 1 dramatically decreases the objective function in the first few number of iterations and the SNR also mainly increases in the first several iterations. Throughtout the following experiments, we set $\lambda=10$ and $N_{ite}=4$.

### *C. Application on Synthetic Seismic Data*

We first apply the HLR approximation method [consists of Algorithm 1 as well as the rank selection scheme in (16)] on a linear-event synthetic data set, which consists of four linear events. It has 300 traces and 500 time-samples per trace. The sampling interval is 2 ms. Fig. 3(a) and (e) shows the clean and noisy data, respectively. To demonstrate our performance, we compare the proposed method with the widely utilized DLR approximation method [3] and the LR approximation method [42], sometimes called the SSA. Similar to [3] and [42], we utilize the block HLR approximation versions of these three methods (LR, DLR, and HLR) when applied for 3-D seismic data. We choose rank $R=2$ for the traditional LR approach and DLR approach. For HLR, we use local processing window size $25\times25\times25$ and utilize the rank selection scheme in (16) to choose the rank for each window. Fig. 3(b)–(d) shows the denoising performance after using the LR, DLR, and HLR methods, respectively. Fig. 3(f)–(h) shows the removed noise that corresponds to the LR, DLR, and HLR methods, respectively. It is obvious that although the conventional LR approach and the DLR approach perform well and remove plenty of random noise, the proposed HLR method does an excellent job by suppressing almost all the random noise. At the same time, the useful signals are not destroyed by all three methods. We note that the range of rank selected by HLR is from 0 to 3 and for most local windows without any event, the HLR can identify the rank as 0 and thus remove the noise. On the other hand, we observe a slight edge effect in Fig. 3(d), which is because the local window processing[1] is adopted in HLR. These edge effects are also observed in Figs. 5 and 6, when LR and DLR are also performed with local windows.

Then, we apply these three methods on a hyperbolic-event 3-D synthetic data set, which is composed of four hyperbolic events in the $x-y-t$ domain. The data contain $N_x\times N_y$ traces with $N_x=300$ and $N_y=300$. The sampling interval is 0.002 s and the total time is 1 s. For the three methods, we use local processing window size $25\times25\times25$, where the hyperbolic events can be considered locally linear. For a seismic profile of the crossline section of the field data example at $N_y=150$, we display the clean and noisy examples in Fig. 4(a) and (b), respectively.

Fig. 4(c) and (d), respectively, shows the denoised data and the removed noise by using the proposed HLR method. It is clear that HLR effectively suppresses most of the random noise, and at the same time, it preserves the useful signals even in the complicated areas. We note that the range of rank selected by HLR is from 0 to 4 and for most local windows without any event, the HLR can identify the rank as 0. Figs. 5 and 6, respectively, describe the denoising performance for the LR method and the DLR method with rank global rank $R$ varying from 1 to 4. We observe from Figs. 5 and 6 that more useful signals are preserved with the rank $R$ increasing from 1 to 4. However, there is also more residual noise in the denoised data when the rank is increasing. It is clear that the useful events are damaged when the rank is 1. Compared with $R=1,2$, the results in $R=3,4$ are better regarding to the protection of useful signals, but are worse with respect to random noise suppression. Compared with the LR approach, the DLR method has a better denoising performance. It seems that when $R=1$, DLR causes the damage of hyperbola events, which is against the denoising principle. When rank is equal to 2, 3, or 4, DLR can suppress plenty of random noise without the loss of useful signals.

We note that the proposed HLR yields an even better performance on random noise reduction. From Fig. 4(c), we see clearly that almost all random noise is eliminated after using the proposed method. In addition, we observe from Fig. 4(d) that there are negligible useful signals in the removed noise by the proposed approach. We also display the SNR obtained by

[1]Local window processing is not adopted in LR and DLR, because they have a better performance when performed globally than locally for linear events.

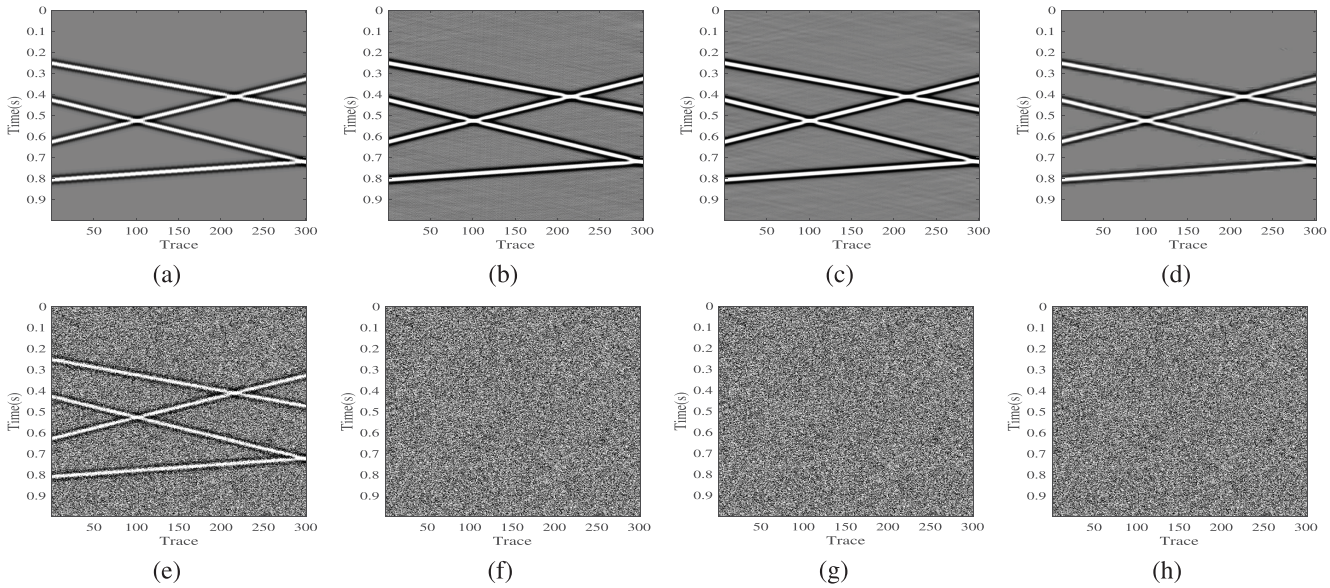


Fig. 3. Denoising comparison of linear-event synthetic data. (a) Clean data and denoised data using (b) LR, (c) DLR, and (d) proposed HLR methods. (e) Noisy data. (f)–(h) Removed noise corresponding to (b)–(d).

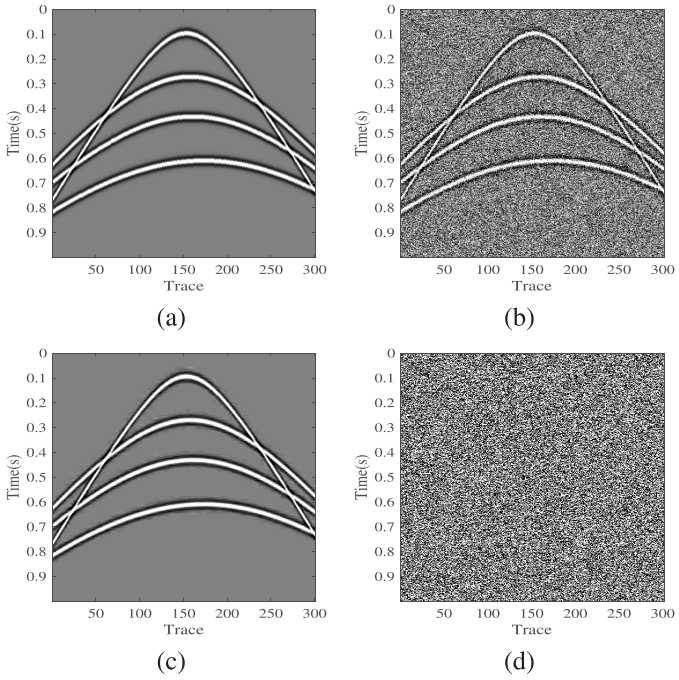


Fig. 4. (a) Clean hyperbolic-event synthetic data. (b) Noisy hyperbolic-event synthetic data. (c) Denoised data using the proposed HLR method. (d) Removed noise by the proposed approach.

different denoising methods in Table I. Note that the proposed HLR is independent of $R$, since it utilizes a novel adaptive rank selection scheme. As can be observed from Table I, the SNR of denoising results by the traditional LR method has a downward trend when the rank $R$ increases from 1 to 4, which means that there is more random noise left with $R$ increasing. When $R = 1$, the LR method obtains a higher SNR of 18.86 dB. The SNRs of denoising results by using the DLR method increase from 23.67 (when $R = 1$) to 28.78 (when $R = 2$), which is followed by a decrease when rank increases from 2 to 4.

TABLE I
COMPARISON OF SNRS FOR DIFFERENT RANK $R$

| $R$ | 1 | 2 | 3 | 4 |
|---|---|---|---|---|
| SNR(dB) of LR | 18.86 | 15.33 | 12.23 | 9.85 |
| SNR(dB) of DLR | 23.67 | 28.78 | 27.27 | 25.35 |
| SNR(dB) of HLR | 31.68 | | | |

TABLE II
COMPARISON OF TIME FOR DIFFERENT METHODS

| | LR | DLR | HLR |
|---|---|---|---|
| Time(s) | 78.06 | 79.12 | 218.35 |

Our proposed method obtains an SNR of 31.68 dB larger than that of any case of traditional LR and DLR methods. Based on the above-mentioned discussions (including all the figures and table), one can draw the conclusion that the proposed HLR yields a superior denoising performance. However, we note that HLR requires more computational time, since it is an iterative method. In Table II, we list the time used for the three methods (all with local window processing) performed by MATLAB 2017 on a Macbook Pro with i5 CPU at 2.3 GHz and RAM 8G. As we explained in Section III, HLR has higher cost, because it requires performing LR approximation in each iteration. A possible method to reduce the computational complexity in HLR is by utilizing the factorization approach [49], [50] that efficiently computes the LR approximation without the computation of SVDs and particularly enables us to use gradient descent for updating $\boldsymbol{L}_k$ in Algorithm 1.

### *D. Application on Field Seismic Data*

To exhibit the superior performance of the proposed method, we apply the HLR method on a real 3-D poststack field

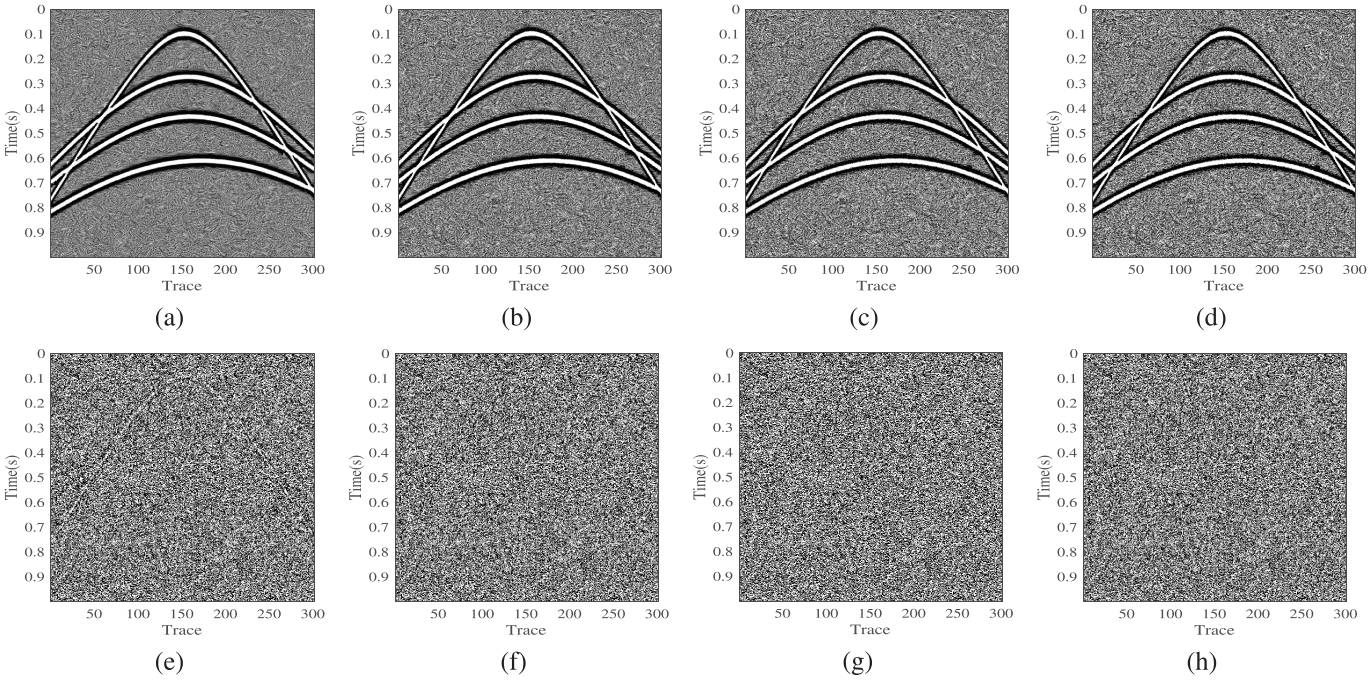


Fig. 5. (a)–(d) Denoised data using the LR method with $R = 1, \ldots, 4$, respectively. (e)–(h) Removed noise corresponding to (a)–(d), respectively.

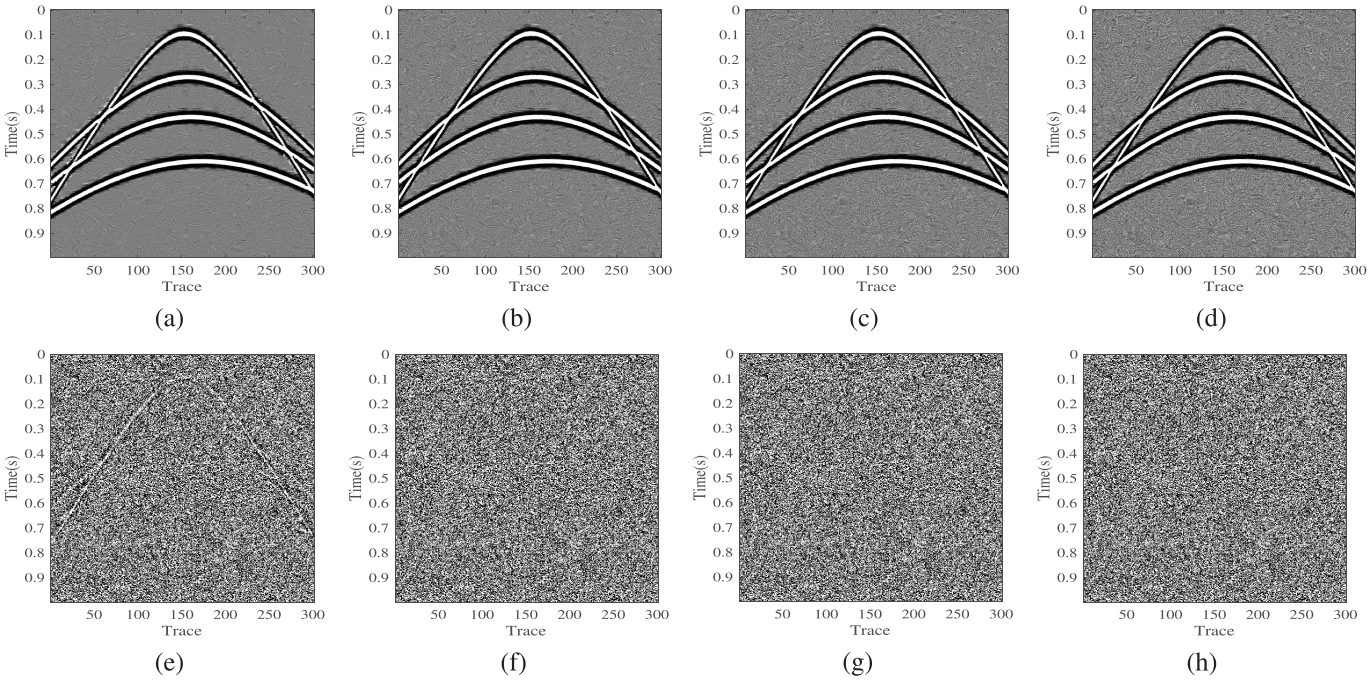


Fig. 6. (a)–(d) Denoised data using the DLR method with $R = 1, \ldots, 4$, respectively. (e)–(h) Removed noise corresponding to (a)–(d), respectively.

data sets. The data contain $N_x \times N_y$ traces with $N_x = 400$ and $N_y = 200$. The sampling interval is 0.004 s. The local processing window size is the same as before, i.e., $25 \times 25 \times 25$. Fig. 7(a) shows the raw noisy seismic data. Fig. 7(b)–(d) shows the corresponding denoised data using the LR method, the DLR method [3], and the proposed HLR method, respectively. Fig. 7(e)–(g) shows the removed noise by these methods. Fig. 8 shows the denoising comparison of the 100th inline section of the field data. Fig. 9 is similar to Fig. 8, but for the 100th crossline section of the field data. We choose $R = 4$ for LR and DLR in order to conservatively preserve the main features of the original data. We can see from Figs. 7–9 that the proposed method achieves a better denoising result with seldom damage of useful signals. From the comparison of the three figures, we also observe that although the denoised results using the LR approach and the DLR approach are quite clean, they lose plenty of useful signals, especially around 0.4–0.6 and 0.6–0.8 s in inline

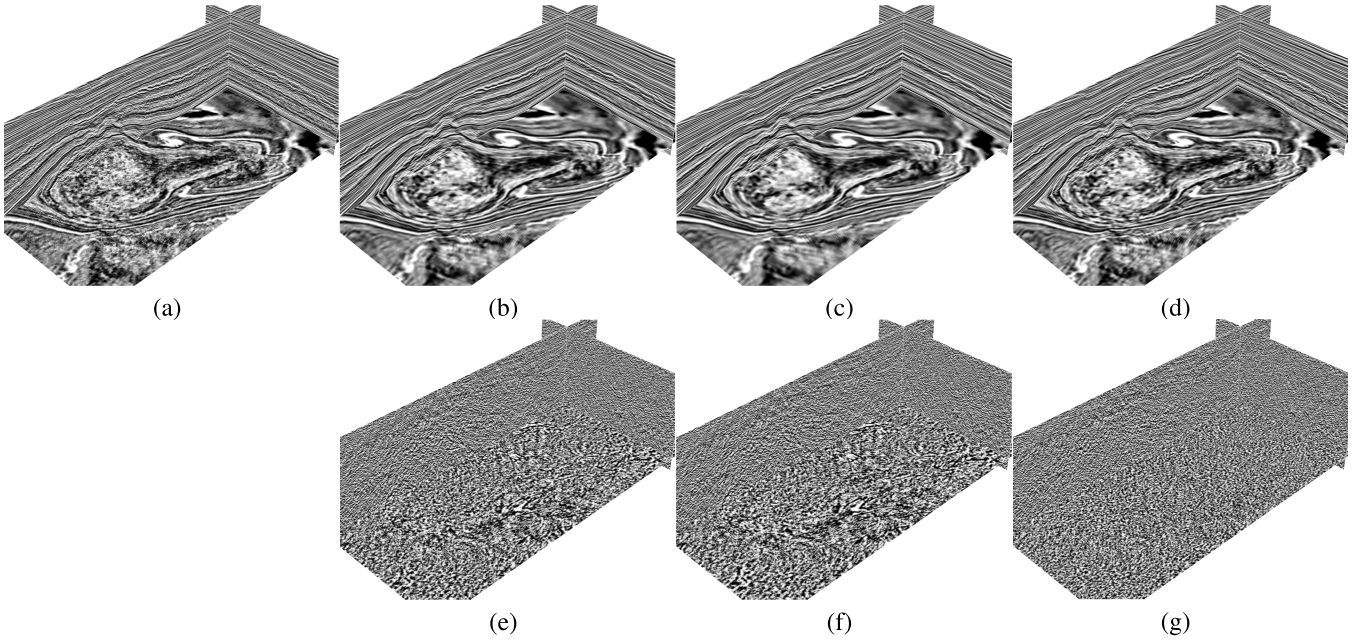


Fig. 7. Denoising comparison of the 3-D field data. (a) Noisy real data and denoised data using (b) LR, (c) DLR, and (d) proposed HLR methods. (e)–(g) Removed noise corresponding to (b)–(d).

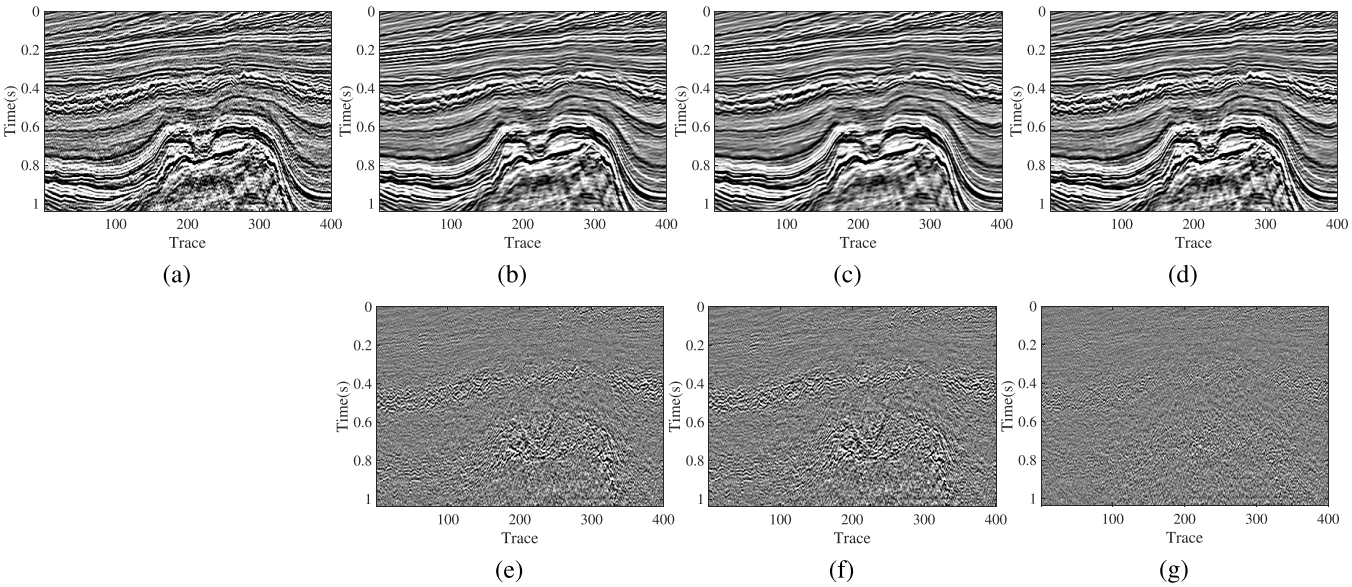


Fig. 8. Denoising comparison of the 100th inline section of the field data. (a) Noisy real data and denoised data using (b) LR, (c) DLR, and (d) proposed HLR methods. (e)–(g) Removed noise corresponding to (b)–(d).

section and 0.4–0.6 and 0.8–1.0 s in crossline section. We can also compare the horizontal slice from Fig. 7(e) and (f) and draw the conclusion that there is obvious damage of useful signals after using the LR approach and the DLR approach. As we already choose $R = 4$ in this test, if we want to preserve as much as useful signal, we need to select a very large rank. But when the rank is too large, the LR and DLR approaches remove seldom random noise. As we can see from Figs. 7(d) and (g), 8(d) and (g), and 9(d) and (g), the proposed HLR can not only remove a significant amount of random noise but also preserve most of the useful signals.

For complex geological structure, there might be several events with different slopes in local window poststack profile, which can be seen clearly in Fig. 8(a). As can be observed from 0.6–0.9 s in Fig. 8(e)–(g), the removed noise by the LR and DLR methods contains an apparent part of the useful signal. While our proposed method produces negligible damage of useful signals. This leads us to the conclusion that the proposed approach can achieve extremely better denoising results, especially in complicated areas.

We finally compare the local correlation coefficients between denoised data and the removed noise. Fig. 10 shows

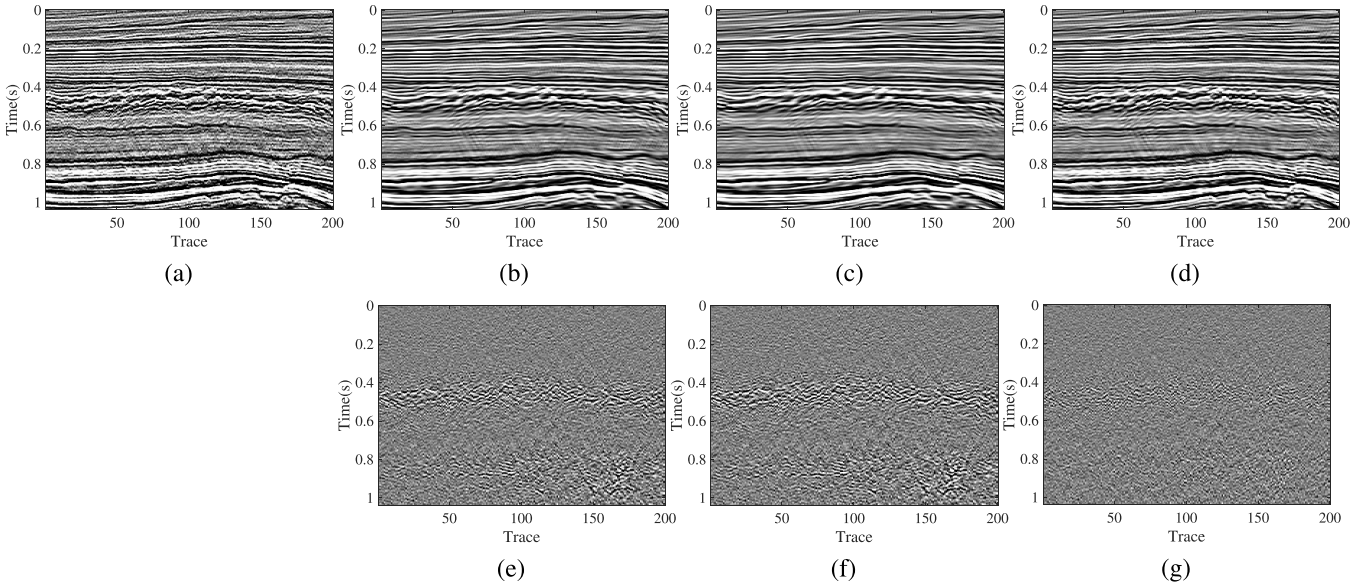


Fig. 9. Denoising comparison of the 100th crossline section of the field data. (a) Noisy real data and denoised data using (b) LR, (c) DLR, and (d) proposed HLR methods. (e)–(g) Removed noise corresponding to (b)–(d).

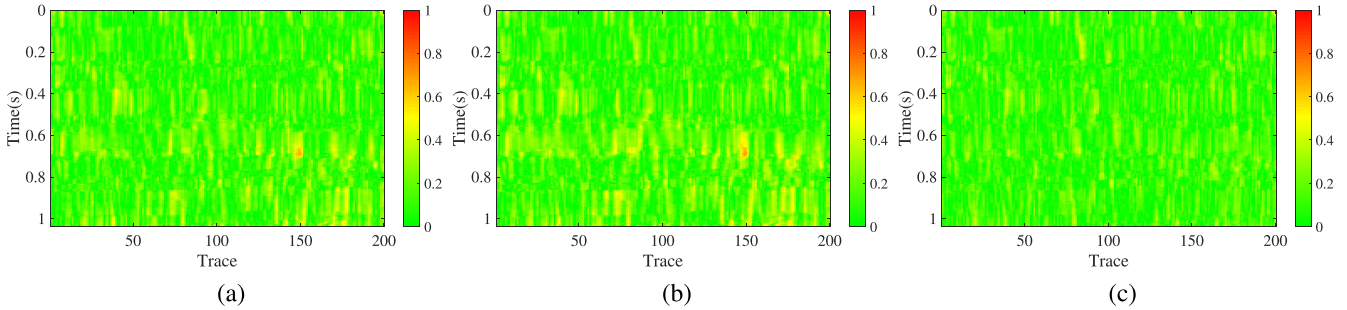


Fig. 10. Local correlation coefficient analysis between denoised data and removed noise at the 100th crossline section of the field data. (a) LR method. (b) DLR method. (c) Proposed HLR method.

the local correlation coefficients of the three methods at the 100th crossline section of the field data. It is clear that there are many high-similarity areas (0.6–0.7 and 0.9–1.0 s) in Fig. 10(a) and (b), revealing the amount of useful signals in the removed noise data. The local correlation coefficient of the HLR method is negligible in most areas, indicating its superior performance in preserving useful signals.

## V. Conclusion

We proposed an HLR scheme to simultaneously leverage both the Hankel structure and the LR property underlying the clean seismic data. The effective denoising performance is improved by exploiting the additional Hankel structure when computing the LR approximation. We propose an alternating-minimization-based algorithm for solving the formulated HLR approximation problem and also provide rigorously convergence analysis of the proposed algorithm. The experiments on both synthetic and field data suggest that the proposed approach achieves superior performance compared with conventional LR approximation methods. In particular, the proposed approach yields extremely better denoising results in complicated areas, such as the local windows consisting of a number of events with different slopes, which widely appear in shot gather or poststack profile. As future work, it is of interest to utilize the factorization approach [49], [50] to efficiently compute the LR approximation without the computation of SVDs.

## Appendix A
## Proof of Theorem 1

We first give out some necessary definitions.

*Definition 1:* Let $f : \mathbb{R}^N \to \mathbb{R} \cup \infty$ be a proper lower semi-continuous function. Its domain, subdifferential, and stationary points are defined as follows.

- The domain of $f$ is defined by $\mathrm{dom} f := \{\boldsymbol{x} \in \mathbb{R}^N : f(\boldsymbol{x}) < \infty\}$.
- For each $\boldsymbol{x} \in \mathrm{dom} f$, $\boldsymbol{x}$ is called the coordinatewise minimum of $f$ if it satisfies

$$f(\boldsymbol{x} + [0, \ldots, d_n, \ldots, 0]^{\mathrm{T}}) \leq f(\boldsymbol{x}), \quad \forall\, d_n, n \in [N].$$

- The subdifferential $\partial f$ is defined by

$$\partial f(\boldsymbol{x}) = \left\{ \boldsymbol{z} : \lim_{\boldsymbol{y} \to \boldsymbol{x}} \inf \frac{f(\boldsymbol{y}) - f(\boldsymbol{x}) - \langle \boldsymbol{z}, \boldsymbol{y} - \boldsymbol{x} \rangle}{\|\boldsymbol{x} - \boldsymbol{y}\|} \geq 0 \right\}$$

for any $\boldsymbol{x} \in \mathrm{dom} f$ and $\partial_F f(\boldsymbol{x}) = \emptyset$ if $\boldsymbol{x} \notin \mathrm{dom} f$.

- For each $\boldsymbol{x} \in \mathrm{dom} f$, $\boldsymbol{x}$ is called the *stationary point* of $f$ if it satisfies $\boldsymbol{0} \in \partial_F f$.

To provide the convergence of Algorithm 1, we first transfer the constrained problem (11) into the the following equivalent form without any constraints:

$$g(\boldsymbol{H}, \boldsymbol{L}) := f(\boldsymbol{H}, \boldsymbol{L}) + \delta_{\mathbb{H}}(\boldsymbol{H}) + \delta_{\mathbb{L}}(\boldsymbol{L}) \tag{21}$$

where $f$ is defined in (11) and $\delta_{\mathbb{H}}(\boldsymbol{H}) = 0$ if $\boldsymbol{H} \in \mathbb{H}$ and $\delta_{\mathbb{H}}(\boldsymbol{H}) = \infty$, if $\boldsymbol{H} \notin \mathbb{H}$, i.e., it is the indicator function of the set $\mathbb{H}$. Similar notation holds for $\delta_{\mathbb{L}}$. We recall that $\mathbb{H}$ and $\mathbb{L}$ are the sets of Hankel matrices and LR matrices, defined in (10) and (14), respectively. We now proceed the proof by individually showing the three arguments in Theorem 1.

### A. Proof (i) of Theorem 1

By the definition of (12), we have

$$\begin{aligned} &f(\boldsymbol{H}_k, \boldsymbol{L}_k) - f(\boldsymbol{H}_{k+1}, \boldsymbol{L}_k) \\ &= (1+\lambda)\left( \left\| \boldsymbol{H}_k - \frac{1}{1+\lambda}(\boldsymbol{D} + \lambda \boldsymbol{L}_k) \right\|_F^2 \right. \\ &\qquad \left. - \left\| (\mathbf{I} - \mathcal{P}_{\mathbb{H}}) \frac{1}{1+\lambda}(\boldsymbol{D} + \lambda \boldsymbol{L}_k) \right\|_F^2 \right) \\ &\geq (1+\lambda) \|\boldsymbol{H}_k - \boldsymbol{H}_{k+1}\|_F^2 \end{aligned} \tag{22}$$

where the last inequality follows because the set $\mathbb{H}$ of Hankel matrices is convex:

$$\langle \boldsymbol{H} - \mathcal{P}_{\mathbb{H}}(\boldsymbol{A}), \mathcal{P}_{\mathbb{H}}(\boldsymbol{A}) - \boldsymbol{A} \rangle \geq 0$$

for any $\boldsymbol{H} \in \mathbb{H}$ and $\boldsymbol{A} \in \mathbb{R}^{M \times N}$. Due to the definition of $\boldsymbol{L}_{k+1}$ in (13), we have

$$f(\boldsymbol{H}_{k+1}, \boldsymbol{L}_{k+1}) \leq f(\boldsymbol{H}_{k+1}, \boldsymbol{L}_k)$$

which together with (22) gives

$$f(\boldsymbol{H}_k, \boldsymbol{L}_k) - f(\boldsymbol{H}_{k+1}, \boldsymbol{L}_{k+1}) \geq (1+\lambda) \|\boldsymbol{H}_k - \boldsymbol{H}_{k+1}\|_F^2. \tag{23}$$

Equation (23) implies that the sequence $\{f(\boldsymbol{H}_k, \boldsymbol{L}_k)\}$ [and also $\{g(\boldsymbol{H}_k, \boldsymbol{L}_k)\}$ by the definition of (21)] is decreasing and hence is convergent by noting that $f$ is nonnegative. Repeating (23) for all $k$ from zero to infinity and adding them up give

$$\sum_{k=0}^{\infty} \|\boldsymbol{H}_k - \boldsymbol{H}_{k+1}\|_F^2 \leq \frac{1}{1+\lambda} f(\boldsymbol{H}_0, \boldsymbol{L}_0)$$

which gives (15).

### B. Proof (ii) of Theorem 1

We first show that

$$\|\boldsymbol{H}_k\|_F, \|\boldsymbol{L}_k\|_F \leq \|\boldsymbol{D}\|_F \tag{24}$$

for all $k$ by induction. To that end, we first assume that $\|\boldsymbol{L}_k\|_F \leq \|\boldsymbol{D}\|_F$. With this, for $\boldsymbol{H}_{k+1}$, we have

$$\begin{aligned} \|\boldsymbol{H}_{k+1}\|_F &= \|\frac{1}{1+\lambda}\mathcal{P}_{\mathbb{H}}(\boldsymbol{D} + \lambda \boldsymbol{L}_k)\|_F \\ &\leq \frac{1}{1+\lambda}\|(\boldsymbol{D} + \lambda \boldsymbol{L}_k)\|_F \\ &\leq \frac{1}{1+\lambda}(\|\boldsymbol{D}\|_F + \lambda \|\boldsymbol{L}_k\|_F) \\ &\leq \|\boldsymbol{D}\|_F \end{aligned}$$

where the first inequality follows from the Cauchy–Schwartz inequality that $\sum_{i=1}^N a_i^2 \geq N(\sum_{i=1}^N a_i/N)^2$ and the last inequality utilizes the fact that $\|\boldsymbol{L}_k\|_F \leq \|\boldsymbol{D}\|_F$.

Similarly, we now assume that $\|\boldsymbol{H}_k\| \leq \|\boldsymbol{D}\|_F$ and show that $\|\boldsymbol{L}_k\|_F \leq \|\boldsymbol{D}\|_F$, which can be obtained by

$$\|\boldsymbol{L}_k\|_F = \|\mathcal{P}_{\mathbb{L}}(\boldsymbol{H}_k)\|_F \leq \|\boldsymbol{H}_k\|_F \leq \|\boldsymbol{D}\|_F.$$

Finally, recall that $\boldsymbol{L}_0 = \mathcal{P}_{\mathbb{L}}(\boldsymbol{D})$, indicating $\|\boldsymbol{L}_0\|_F \leq \|\boldsymbol{D}\|_F$. This completes the proof for (24).

For convenience, throughout the proof, we denote by $\boldsymbol{Z}_k = \{(\boldsymbol{H}_k, \boldsymbol{L}_k)\}$. Since we have shown that the sequence $\{(\boldsymbol{Z}_k)\}$ is bounded, the Bolzano–Weiestrass theorem [51] states that this sequence has at least one convergent subsequence. Let $\boldsymbol{Z}^\star = (\boldsymbol{H}^\star, \boldsymbol{L}^\star)$ be the limit point of any convergent subsequence $\{\boldsymbol{Z}_{k_m}\}$.

Since $\mathbb{H}$ is convex, we have $\boldsymbol{H}^\star \in \mathbb{H}$ and hence

$$\lim_{k_m \to \infty} \delta_{\mathbb{H}}(\boldsymbol{H}_{k_m}) = \delta_{\mathbb{H}}(\boldsymbol{H}^\star) = 0.$$

Due to the fact that $f$ is smooth, we also have

$$\lim_{k_m \to \infty} f(\boldsymbol{Z}_{k_m}) = f(\boldsymbol{Z}^\star).$$

By the definition of (13), we have

$$f(\boldsymbol{H}_{k_m}, \boldsymbol{L}_{k_m}) + \delta_{\mathbb{L}}(\boldsymbol{L}_{k_m}) \leq f(\boldsymbol{H}_{k_m}, \boldsymbol{L}) + \delta_{\mathbb{L}}(\boldsymbol{L}) \tag{25}$$

for all $\boldsymbol{L} \in \mathbb{R}^{M \times N}$. Taking $k_m \to \infty$ and substituting $\boldsymbol{L}$ by $\boldsymbol{L}^\star$ into (25), and invoking the fact that $\delta_{\mathbb{L}}(\boldsymbol{L})$ is lower semicontinuous, we have

$$\liminf_{k_m \to \infty} \delta_{\mathbb{H}}(\boldsymbol{L}_{k_m}) = \delta_{\mathbb{H}}(\boldsymbol{L}^\star)$$

which together with the fact that the sequence $\{g(\boldsymbol{Z}_k)\}$ is convergent gives

$$\lim_{k_m \to \infty} \delta_{\mathbb{H}}(\boldsymbol{L}_{k_m}) = \delta_{\mathbb{H}}(\boldsymbol{L}^\star). \tag{26}$$

Therefore, we have

$$\begin{aligned} &\lim_{k_m \to \infty} g(\boldsymbol{Z}_{k_m}) \\ &\quad = \lim_{k_m \to \infty} f(\boldsymbol{Z}_{k_m}) + \lim_{k_m \to \infty} \delta_{\mathbb{H}}(\boldsymbol{H}_{k_m}) + \lim_{k_m \to \infty} \delta_{\mathbb{H}}(\boldsymbol{L}_{k_m}) \\ &\quad = g(\boldsymbol{Z}^\star). \end{aligned}$$

Finally, because $\{g(\boldsymbol{Z}_k)\}$ is decreasing and $g(\boldsymbol{Z}_k) \geq 0$, we have

$$f(\boldsymbol{Z}^\star) = g(\boldsymbol{Z}^\star) = \inf_k g(\boldsymbol{Z}_k) = \inf_k f(\boldsymbol{Z}_k).$$

### C. Proof (iii) of Theorem 1

We first show that $\boldsymbol{Z}^\star = (\boldsymbol{H}^\star, \boldsymbol{L}^\star)$ is a coordinatewise minimum [see (1)]. Letting $k_m \to \infty$ in (25) and invoking (26), we have

$$f(\boldsymbol{Z}^\star) + \delta_{\mathbb{H}}(\boldsymbol{H}^\star) + \delta_{\mathbb{L}}(\boldsymbol{L}^\star) \le f(\boldsymbol{H}^\star, \boldsymbol{Z}) + \delta_{\mathbb{H}}(\boldsymbol{H}^\star) + \delta_{\mathbb{L}}(\boldsymbol{L})$$

for all $\boldsymbol{L} \in \mathbb{R}^{M\times N}$, which further implies

$$g(\boldsymbol{Z}^\star) \le g(\boldsymbol{H}^\star, \boldsymbol{L}^\star + \boldsymbol{L}), \quad \forall \boldsymbol{L} \in \mathbb{R}^{M\times N}. \tag{27}$$

With similar argument, we also have

$$g(\boldsymbol{Z}^\star) \le g(\boldsymbol{H}^\star + \boldsymbol{H}, \boldsymbol{L}^\star), \quad \forall \boldsymbol{H} \in \mathbb{R}^{M\times N}. \tag{28}$$

Therefore, $\boldsymbol{Z}^\star$ is a coordinatewise minimum.

The remaining part is to show that $\boldsymbol{Z}^\star$ is a stationary point, which, by (1), is equivalent to show

$$\liminf_{\|\boldsymbol{\Delta}_{\boldsymbol{Z}}\|\to 0} \frac{g(\boldsymbol{Z}^\star + \boldsymbol{\Delta}_{\boldsymbol{Z}}) - g(\boldsymbol{Z}^\star)}{\|\boldsymbol{\Delta}_{\boldsymbol{Z}}\|} \ge 0 \tag{29}$$

where $\boldsymbol{\Delta}_{\boldsymbol{Z}} = (\boldsymbol{\Delta}_{\boldsymbol{H}}, \boldsymbol{\Delta}_{\boldsymbol{L}})$ with $\boldsymbol{\Delta}_{\boldsymbol{H}}, \boldsymbol{\Delta}_{\boldsymbol{L}} \in \mathbb{R}^{M\times N}$. To prove (29), we first utilize Taylor's theorem to expand $f(\boldsymbol{Z}^\star + \boldsymbol{\Delta}_{\boldsymbol{Z}})$ at $\boldsymbol{Z}^\star$ when $\|\boldsymbol{\Delta}_{\boldsymbol{Z}}\|_F \to 0$

$$\begin{aligned} f(\boldsymbol{Z}^\star + \boldsymbol{\Delta}_{\boldsymbol{Z}}) - f(\boldsymbol{Z}^\star) &= \langle \nabla f(\boldsymbol{Z}^\star), \boldsymbol{\Delta}_{\boldsymbol{Z}} \rangle + o(\|\boldsymbol{\Delta}_{\boldsymbol{Z}}\|) \\ &= f(\boldsymbol{H}^\star + \boldsymbol{\Delta}_{\boldsymbol{H}}, \boldsymbol{L}^\star) - f(\boldsymbol{Z}^\star) \\ &\quad + f(\boldsymbol{H}^\star, \boldsymbol{L}^\star + \boldsymbol{\Delta}_{\boldsymbol{L}}) - f(\boldsymbol{Z}^\star) \\ &\quad + o(\|\boldsymbol{\Delta}_{\boldsymbol{Z}}\|) \end{aligned}$$

where the "little-o" notation $o(c)$ is defined by $\lim_{c\to 0} o(c)/c = 0$. With the above equation, we now prove (29) as

$$\begin{aligned} &\liminf_{\|\boldsymbol{\Delta}_{\boldsymbol{Z}}\|_F\to 0} \frac{g(\boldsymbol{Z}^\star + \boldsymbol{\Delta}_{\boldsymbol{Z}}) - g(\boldsymbol{Z}^\star)}{\|\boldsymbol{\Delta}_{\boldsymbol{Z}}\|_F} \\ &\quad = \liminf_{\|\boldsymbol{\Delta}_{\boldsymbol{Z}}\|_F\to 0} \left( \frac{g(\boldsymbol{H}^\star + \boldsymbol{\Delta}_{\boldsymbol{H}}, \boldsymbol{L}^\star) - g(\boldsymbol{Z}^\star)}{\|\boldsymbol{\Delta}_{\boldsymbol{Z}}\|_F} \right. \\ &\qquad \left. + \frac{g(\boldsymbol{H}^\star, \boldsymbol{L}^\star + \boldsymbol{\Delta}_{\boldsymbol{L}}) - g(\boldsymbol{Z}^\star)}{\|\boldsymbol{\Delta}_{\boldsymbol{Z}}\|_F} \right) \ge 0 \end{aligned}$$

where the last inequality follows from (27) and (28).

This completes the proof of Theorem 1.

## ACKNOWLEDGMENTS

The authors would like to thank the associate editor and reviewers for their helpful comments and constructive suggestions that help improve the quality of this paper.

## REFERENCES

[1] J. M. Chen and J. Cihlar, "Quantifying the effect of canopy architecture on optical measurements of leaf area index using two gap size analysis methods," *IEEE Trans. Geosci. Remote Sens.*, vol. 33, no. 3, pp. 777–787, May 1995.

[2] Y. Chen, D. Zhang, W. Huang, and W. Chen, "An open-source MATLAB code package for improved rank-reduction 3D seismic data denoising and reconstruction," *Comput. Geosci.*, vol. 95, pp. 59–66, Oct. 2016.

[3] Y. Chen *et al.*, "Simultaneous denoising and reconstruction of 5-D seismic data via damped rank-reduction method," *Geophys. J. Int.*, vol. 206, no. 3, pp. 1695–1717, 2016.

[4] X. Wu and D. Hale, "3D seismic image processing for faults," *Geophysics*, vol. 81, no. 2, pp. IM1–IM11, 2016.

[5] X. Wu, "Directional structure-tensor-based coherence to detect seismic faults and channels," *Geophysics*, vol. 82, no. 2, pp. A13–A17, 2017.

[6] X. Wu and Z. Zhu, "Methods to enhance seismic faults and construct fault surfaces," *Comput. Geosci.*, vol. 107, pp. 37–48, Oct. 2017.

[7] F. Li, H. Zhou, T. Zhao, and K. Marfurt, "Unconventional reservoir characterization based on spectrally corrected seismic attenuation estimation," *J. Seismic Exploration*, vol. 25, no. 5, pp. 447–461, 2016.

[8] X. Wu, "Structure-, stratigraphy- and fault-guided regularization in geophysical inversion," *Geophys. J. Int.*, vol. 210, no. 1, pp. 184–195, 2017.

[9] H. Krim, D. Tucker, S. Mallat, and D. Donoho, "On denoising and best signal representation," *IEEE Trans. Inf. Theory*, vol. 45, no. 7, pp. 2225–2238, Nov. 1999.

[10] Z. Zhu and M. B. Wakin, "Approximating sampled sinusoids and multiband signals using multiband modulated DPSS dictionaries," *J. Fourier Anal. Appl.*, vol. 23, no. 6, pp. 1263–1310, Dec. 2017.

[11] E. J. Candès and M. B. Wakin, "An introduction to compressive sampling," *IEEE Signal Process. Mag.*, vol. 25, no. 2, pp. 21–30, Mar. 2008.

[12] Z. Zhu, G. Li, J. Ding, Q. Li, and X. He, "On collaborative compressive sensing systems: The framework, design, and algorithm," *SIAM J. Imag. Sci.*, vol. 11, no. 2, pp. 1717–1758, 2018.

[13] M. D. Sacchi, T. J. Ulrych, and C. J. Walker, "Interpolation and extrapolation using a high-resolution discrete Fourier transform," *IEEE Trans. Signal Process.*, vol. 46, no. 1, pp. 31–38, Jan. 1998.

[14] D. L. Donoho and J. M. Johnstone, "Ideal spatial adaptation by wavelet shrinkage," *Biometrika*, vol. 81, no. 3, pp. 425–455, 1994.

[15] F. J. Herrmann and G. Hennenfent, "Non-parametric seismic data recovery with curvelet frames," *Geophys. J. Int.*, vol. 173, no. 1, pp. 233–248, 2008.

[16] T. Qiao *et al.*, "Effective denoising and classification of hyperspectral images using curvelet transform and singular spectrum analysis," *IEEE Trans. Geosci. Remote Sens.*, vol. 55, no. 1, pp. 119–133, Jan. 2017.

[17] H. Shan, J. Ma, and H. Yang, "Comparisons of wavelets, contourlets and curvelets in seismic denoising," *J. Appl. Geophys.*, vol. 69, no. 2, pp. 103–115, 2009.

[18] S. Liu, M. Liu, P. Li, J. Zhao, Z. Zhu, and X. Wang, "SAR image denoising via sparse representation in shearlet domain based on continuous cycle spinning," *IEEE Trans. Geosci. Remote Sens.*, vol. 55, no. 5, pp. 2985–2992, May 2017.

[19] Y. Chen, S. Fomel, and J. Hu, "Iterative deblending of simultaneous-source seismic data using seislet-domain shaping regularization," *Geophysics*, vol. 79, no. 5, pp. V179–V189, 2014.

[20] Y. Chen and S. Fomel, "EMD-seislet transform," *Geophysics*, vol. 83, no. 1, pp. A27–A32, 2018.

[21] D. O. Trad, T. J. Ulrych, and M. D. Sacchi, "Accurate interpolation with high-resolution time-variant Radon transforms," *Geophysics*, vol. 67, no. 2, pp. 644–656, 2002.

[22] Y. Chen, J. Ma, and S. Fomel, "Double-sparsity dictionary for seismic noise attenuation," *Geophysics*, vol. 81, no. 2, pp. V103–V116, 2016.

[23] Y. Chen, "Fast dictionary learning for noise attenuation of multidimensional seismic data," *Geophys. J. Int.*, vol. 209, no. 1, pp. 21–31, 2017.

[24] M. A. N. Siahsar, S. Gholtashi, V. Abolghasemi, and Y. Chen, "Simultaneous denoising and interpolation of 2D seismic data using data-driven non-negative dictionary learning," *Signal Process.*, vol. 141, pp. 309–321, Dec. 2017.

[25] Y. Li, Y. Zhang, X. Huang, H. Zhu, and J. Ma, "Large-scale remote sensing image retrieval by deep hashing neural networks," *IEEE Trans. Geosci. Remote Sens.*, vol. 56, no. 2, pp. 950–965, Feb. 2018.

[26] Y. LeCun, Y. Bengio, and G. Hinton, "Deep learning," *Nature*, vol. 521, no. 7553, pp. 436–444, 2015.

[27] Y. Li, Y. Zhang, X. Huang, and J. Ma, "Learning source-invariant deep hashing convolutional neural networks for cross-source remote sensing image retrieval," *IEEE Trans. Geosci. Remote Sens.*, to be published, doi: 10.1109/TGRS.2018.2839705.

[28] R. Abma and J. Claerbout, "Lateral prediction for noise attenuation by *t-x* and *f-x* techniques," *Geophysics*, vol. 60, no. 6, pp. 1887–1896, 1995.

[29] N. Gulunay *et al.*, "FXDECON and complex wiener prediction filter," in *Proc. SEG Annu. Meeting Soc. Exploration Geophysici.*, 1986, pp. 279–281.

[30] G.-C. Liu, X.-H. Chen, J. Du, and J.-W. Song, "Seismic noise attenuation using nonstationary polynomial fitting," *Appl. Geophys.*, vol. 8, no. 1, pp. 18–26, 2011.

[31] G. Liu, X. Chen, J. Du, and K. Wu, "Random noise attenuation using *f-x* regularized nonstationary autoregression," *Geophysics*, vol. 77, no. 2, pp. V61–V69, 2012.

[32] Y. Kopsinis and S. McLaughlin, "Development of EMD-based denoising methods inspired by wavelet thresholding," *IEEE Trans. Signal Process.*, vol. 57, no. 4, pp. 1351–1362, Apr. 2009.
[33] Y. Chen and J. Ma, "Random noise attenuation by *f-x* empirical-mode decomposition predictive filtering," *Geophysics*, vol. 79, no. 3, pp. V81–V91, 2014.
[34] S. Yu and J. Ma, "Complex variational mode decomposition for slop-preserving denoising," *IEEE Trans. Geosci. Remote Sens.*, vol. 56, no. 1, pp. 586–597, Jan. 2018.
[35] Y. Chen, "Dip-separated structural filtering using seislet transform and adaptive empirical mode decomposition based dip filter," *Geophys. J. Int.*, vol. 206, no. 1, pp. 457–469, 2016.
[36] N. Kreimer and M. D. Sacchi, "A tensor higher-order singular value decomposition for prestack seismic data noise reduction and interpolation," *Geophysics*, vol. 77, no. 3, pp. V113–V122, 2012.
[37] Y. Wang, H. Zhou, S. Zu, W. Mao, and Y. Chen, "Three-operator proximal splitting scheme for 3-D seismic data reconstruction," *IEEE Geosci. Remote Sens. Lett.*, vol. 14, no. 10, pp. 1830–1834, Oct. 2017.
[38] Y. Chen and S. Fomel, "Random noise attenuation using local signal-and-noise orthogonalization," *Geophysics*, vol. 80, no. 6, pp. WD1–WD9, 2015.
[39] Y. Chen, "Non-stationary least-squares complex decomposition for microseismic noise attenuation," *Geophys. J. Int.*, vol. 213, no. 3, pp. 1572–1585, 2018.
[40] J. A. Cadzow, "Signal enhancement-a composite property mapping algorithm," *IEEE Trans. Acoust., Speech, Signal Process.*, vol. 36, no. 1, pp. 49–62, Jan. 1988.
[41] R. Vautard and M. Ghil, "Singular spectrum analysis in nonlinear dynamics, with applications to paleoclimatic time series," *Phys. D, Nonlinear Phenom.*, vol. 35, no. 3, pp. 395–424, 1989.
[42] V. Oropeza and M. Sacchi, "Simultaneous seismic data denoising and reconstruction via multichannel singular spectrum analysis," *Geophysics*, vol. 76, no. 3, pp. V25–V32, 2011.
[43] M. A. N. Siahsar, S. Gholtashi, E. O. Torshizi, W. Chen, and Y. Chen, "Simultaneous denoising and interpolation of 3-D seismic data via damped data-driven optimal singular value shrinkage," *IEEE Geosci. Remote Sens. Lett.*, vol. 14, no. 7, pp. 1086–1090, Jul. 2017.
[44] Y. Chen, Y. Zhou, W. Chen, S. Zu, W. Huang, and D. Zhang, "Empirical low-rank approximation for seismic noise attenuation," *IEEE Trans. Geosci. Remote Sens.*, vol. 55, no. 8, pp. 4696–4711, Aug. 2017.
[45] J. Sun, S. Fomel, T. Zhu, and J. Hu, "Q-compensated least-squares reverse time migration using low-rank one-step wave extrapolation," *Geophysics*, vol. 81, no. 4, pp. S271–S279, 2016.
[46] H. Attouch, J. Bolte, P. Redont, and A. Soubeyran, "Proximal alternating minimization and projection methods for nonconvex problems: An approach based on the Kurdyka-Łojasiewicz inequality," *Math. Oper. Res.*, vol. 35, no. 2, pp. 438–457, 2010.
[47] Z. Zhu and X. Li. (2018). "Convergence analysis of alternating nonconvex projections." [Online]. Available: https://arxiv.org/abs/1802.03889
[48] M. Gavish and D. L. Donoho, "The optimal hard threshold for singular values is $4/\sqrt{3}$," *IEEE Trans. Inf. Theory*, vol. 60, no. 8, pp. 5040–5053, Jun. 2014.
[49] Q. Li, Z. Zhu, and G. Tang, "The non-convex geometry of low-rank matrix optimization," *Inf. Inference A, J. IMA*, to be published, doi: 10.1093/imaiai/iay003.
[50] Z. Zhu, Q. Li, G. Tang, and M. B. Wakin, "Global optimality in low-rank matrix optimization," *IEEE Trans. Signal Process.*, vol. 66, no. 13, pp. 3614–3628, Jul. 2018.
[51] R. G. Bartle and D. R. Sherbert, *Introduction to Real Analysis*, vol. 2. New York, NY, USA: Wiley, 2000.

**Chong Wang** is currently pursuing the Ph.D. degree with the Institute of Geophysics and Geomatics, China University of Geosciences, Wuhan, China.

She is also a Visiting Scholar with the Bureau of Economic Geology, The University of Texas at Austin, Austin, TX, USA. Her research interests include seismic data processing and inversion.

**Zhihui Zhu** (S'17) received the B.Eng. degree in communications engineering from the Zhejiang University of Technology, Hangzhou, China, in 2012, and the Ph.D. degree in electrical engineering from the Colorado School of Mines, Golden, CO, USA, in 2017.

He is currently a Post-Doctoral Fellow with the Center for Imaging Science, Johns Hopkins University, Baltimore, MD, USA. His research interests include exploiting inherent structures and applying optimization methods with guaranteed performance for signal processing, machine learning, and data analysis.

Dr. Zhu was the Meritorious Winner of the Mathematical Contest in Modeling/Interdisciplinary Contest in Modeling in 2011.

**Hanming Gu** received the B.S., M.S., and Ph.D. degrees in geophysical detection and information technology with an emphasis on geophysics from the China University of Geosciences, Wuhan, China, in 1985, 1988, and 2000, respectively.

He is currently a Professor with the China University of Geosciences. His research interests include seismic acquisition, seismic processing, and modeling in reservoir geophysics.

**Xinming Wu** received the Ph.D. degree in geophysics from the Colorado School of Mines, Golden, CO, USA, in 2016.

He was a member of the Center for Wave Phenomena, Colorado School of Mines. He is currently a Post-Doctoral Fellow with The University of Texas at Austin, Austin, TX, USA. His research interests include image processing, 3-D seismic interpretation, subsurface modeling, and geophysical inversion.

Dr. Wu received the awards for the Best Paper in Geophysics in 2016 and the Best Student Poster Paper presented at the 2017 SEG Annual Meeting.

**Shuaiqi Liu** received the B.S. degree from the Department of Information and Computer Science, Shandong University of Science and Technology, Qingdao, China, in 2009, and the Ph.D. degree from the Institute of Information Science, Beijing Jiaotong University, Beijing, China, in 2014.

He was a Visiting Scholar at Ottawa University, Ottawa, ON, Canada, from 2016 to 2017. He is currently an Associate Professor with the College of Electronic and Information Engineering, Hebei University, Baoding, China. His research interests include image processing and signal processing.